Electronic supplementary material

Paper title

A molecular timescale for eukaryote evolution recalibrated with the continuous microfossil record

Authors

Cédric Berney [1] and Jan Pawlowski *

Department of Zoology and Animal Biology, University of Geneva, Sciences III,
30, quai Ernest Ansermet, 1211 Geneva 4, Switzerland

* Corresponding author: jan.pawlowski@zoo.unige.ch

[1] Present address:
Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

1) Methods :

Additional details on the phylogenetic and dating analyses, and on our choice of four maximum time constraints.

Use of outgroup sequences in our phylogenetic and dating analyses

Outgroup sequences are necessary in the first steps of the dating analysis in order to root the phylogeny of the ingroup. However, they are automatically removed in the last step (the actual dating process) because there is no way for the program to determine where the root should be placed on the branch separating the outgroup from the ingroup. Therefore, the earlier node to which a date can be assigned is the first node at the base of the ingroup. In our case, we decided to conduct the initial phylogenetic analyses in absence of prokaryotic outgroup sequences, because they are so distant from the eukaryotic sequences in the case of SSU rDNA that they systematically lead to long-branch attraction (LBA) artefacts in tree reconstructions. We could thus select 1465 unambiguously aligned positions to use in the phylogenetic analyses. As we also excluded particularly fast-evolving eukaryotic lineages from the dataset, we are confident that the topology we obtained was not significantly biased by LBA artefacts. Because we wanted to be able to (1) get an estimate of the date at the base of the radiation of extant eukaryotes, and (2) test several possible positions for the root of the eukaryotic tree (see supplementary table 2), we added two archaebacterial sequences (*Sulfolobus acidocaldarius* and *Thermococcus celer*; GenBank accession numbers D14876 and M21529, respectively) to our dataset of 83 eukaryotic sequences for the dating analysis. To do this, we aligned those regions that are unambiguously homologous between eukaryotes and Archaea, and in the two archaebacterial sequences, we coded as missing data the regions for which homology with eukaryotic sequences could not be assumed. These archaebacterial sequences were used to artificially constrain four possible positions for the root of the eukaryote tree, and were automatically pruned during the final step of the dating analysis. The four possible positions for the root of the eukaryote tree were tested by constraining the position of the two outgroup sequences in the best tree topology obtained without outgroup: (1) on the branch separating unikonts from bikonts, (2) on the branch separating opisthokonts from all other eukaryotes, (3) on the branch separating Amoebozoa from all other eukaryotes, and (4) on the branch separating Excavates from all other eukaryotes.

Prior gamma distributions on the parameters of the relaxed clock model

Prior gamma distributions on three parameters of the relaxed clock model were assumed and specified through the mean and standard deviation (SD) of the root age, root rate, and rate autocorrelation. At least three different plausible priors on the root age (*a priori* expected time between tips and root) were tested for each analysis until convergence of the dating procedure to an estimated date at the root corresponding to the prior. Other priors and their SDs were set to the values recommended by the authors (Kishino *et al*. 2001). Namely, the SD for the prior on the root age was set at half the prior on the root age; the prior on the root rate (nucleotide replacements per 100 sites per My at the ingroup root node) was given as the median sum of branch lengths between tips and root divided by the prior on the root age; the SD for the prior on the root rate was set at half the prior on the root rate; and finally, both the prior on the parameter that controls the degree of rate autocorrelation per My along the descending branches of the tree (ν) and its SD were set at 1 / the prior on the root age. The highest possible time between tips and root ("bigtime") was arbitrarily set at 4500 Mya for all analyses.

*Selecting the four maximum time constraints used as calibration points*

Our first MaxTC is the radiation of extant peridinin-containing dinoflagellate lineages that took place during the Triassic. Recent evidence from both the pattern of the fossil record (Fensome *et al*. 1996) and the record of dinosteroids (e.g. Moldowan *et al*. 1996) support a rapid radiation of these organisms at the beginning of the Mesozoic, and exclude the hypothesis that the sudden appearance and diversification of dinoflagellates in the fossil record was preceded by a long period of evolution of non-fossilizable species (Fensome *et al*. 1996). We used a conservative MaxTC of 250 My for the age of this radiation. The second MaxTC was provided by the fossil record of coccolithophorids. Although coccoliths are documented in sediments of up to 300 My old (Siesser 1993), all modern coccolithophorids belong to families that developed after the Cretaceous-Tertiary boundary (Young *et al*. 1994). In a conservative approach allowing the possible divergence of these families shortly before 65 Mya, a MaxTC of 65 My was set at the radiation of one of the two extant main lineages of coccolithophorids. Our last two MaxTCs concern diatoms, which display very well-documented fossil records in the mid-Cretaceous and in the Tertiary. The exact time of appearance of the pennate diatoms is unknown because of a period of poor silica deposition in the upper Cretaceous. However, they are abundant in the Tertiary and completely absent in the mid-Cretaceous, 110 Mya (Kooistra & Medlin 1996). In order to take into account this gap in the diatoms fossil record, we conservatively chose the date of 110 My as a MaxTC for the divergence of pennate diatoms from their centric ancestor. Our fourth MaxTC relates to the radiation of the rhizosolenid diatoms. Although the time of this radiation cannot be directly deduced from the fossil record (because it also occurred in the above-mentioned period of poor silica deposition in the upper Cretaceous), a precise estimate of this radiation can be indirectly inferred from the presence of specific biomarkers of the rhizosolenids in sediments of up to 90 My old (Sinninghe Damsté *et al*. 2004), which provided our second diatom MaxTC.

Additional references not cited in the main text :

Fensome, R. A., MacRae, R. A., Moldowan, J. M., Taylor, F. J. R. & Williams, G. L. 1996 The early Mesozoic radiation of dinoflagellates. *Paleobiology* **22**, 329–338.

Moldowan, J. M., Dahl, J., Jacobson, S. R., Huizinga, B. J., Fago, F. J., Shetty, R., Watt, D. S. & Peters, K. E. 1996 Chemostratigraphic reconstruction of biofacies: molecular evidence linking cyst-forming dinoflagellates with pre-Triassic ancestors. *Geology* **24**, 159–162.

Siesser, W. G. 1993 Calcareous nannoplankton. In *Fossil Prokaryotes and Protists* (ed. J. H. Lipps), pp. 169–201. Boston: Blackwell Scientific Publications.

Young, J., Brown, P. R. & Burnett, L. A. 1994 Palaeontological perspectives. In *The haptophyte algae* (ed. J. C. Green & B. S. C. Leadbeater), pp. 379–392. Oxford: Systematic Association special volume 51, Clarendon Press.

Kooistra, W. H. C. F. & Medlin, L. K. 1996 Evolution of the diatoms (Bacillariophyta) IV. A reconstruction of their age from small subunit rRNA coding regions and the fossil record. *Mol. Phylogenet. Evol.* **6**, 391–407.

Sinninghe Damsté, J. S., Muyzer, G., Abbas, B., Rampen, S. W., Massé, G., Allard, W. G., Belt, S. T., Robert, J. M., Rowland, S. J., Moldowan, J. M., Barbanti, S. M., Fago, F. J., Denisevich, P., Dahl, J., Trindade, L. A. F. & Schouten, S. 2004 The rise of rhizosolenid diatoms. *Science* **304**, 584–587.

2) Supplementary Figure 1.

Maximum likelihood (left) and Bayesian (right) phylograms. All branches are drawn to scale. Numbers at nodes (Bayesian tree) are the posterior probabilities.
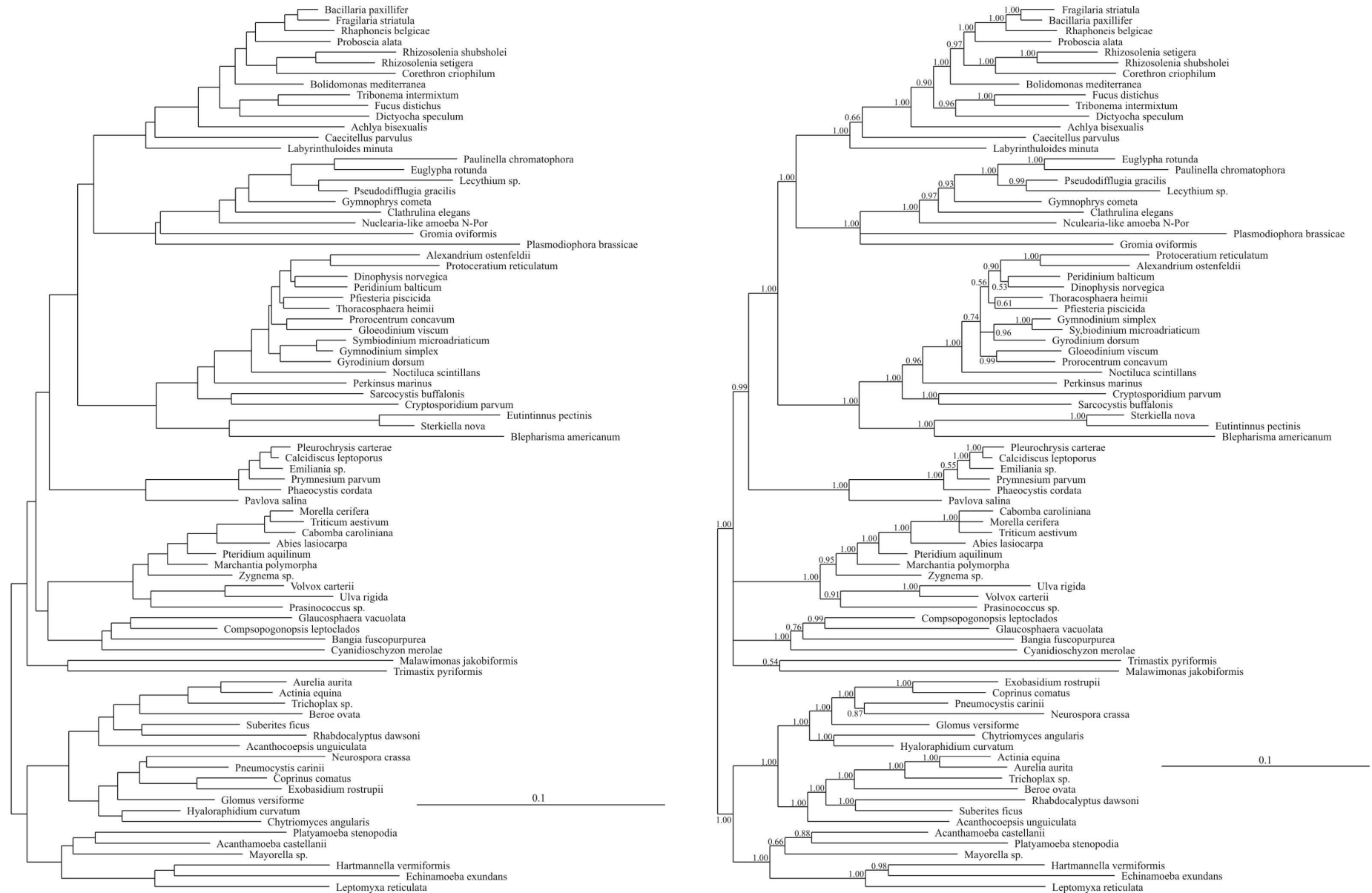
3) Supplementary Table 1. Date and description of all fossil events used as calibration points or discussed in this study, along with their references.

Taxon	Fossil event	Date and/or calibration point [a]	References
Eukaryotes	presence of steranes (eukaryote-specific biomarkers?)	~ 2500-2800	Brocks et al. 1999 ; Brocks et al. 2003
	Grypania sp.	~ 1800-2100	Kumar 1995 ; Han & Runnegar 1992
Amoebozoa	*Prantlitina* sp. (a difflugiid, testate amoeba ?)	325	Loeblich & Tappan 1964
	Vase-shaped microfossils interpreted as testate, lobose amoebae	L 750 (node 6)	Porter & Knoll 2000 ; Porter et al. 2003
Alveolates	Range of the family Goniodomaceae	L 140 (node 60)	Fensome et al. 1996
	Range of the family Gonyaulacaceae	L 200 (node 58)	Fensome et al. 1996
	Radiation of extant crown dinoflagellate lineages in the Triassic	L 210 (node 51)	Fensome et al. 1996 ; Summons et al. 1992
	Radiation of extant crown dinoflagellate lineages in the Triassic	U 250 (node 51)	Fensome et al. 1996 ; Summons et al. 1992
	Oldest known undisputable crown dinoflagellate	L 240 (node 50)	Fensome et al. 1996
Haptophytes	oldest known coccolithophorids	210-300	Siesser 1993
	Range of the genus *Calcidiscus*	L 23 (node 42)	Sáez et al. 2003
	Post K/T radiation of extant coccolithophorids	U 65 (node 42)	Young et al. 1994 ; Fujiwara et al. 2001
	Post K/T radiation of extant coccolithophorids	L 65 (node 40)	Young et al. 1994 ; Fujiwara et al. 2001
Opisthokonts	Oldest known undisputable members of extant bilaterian phyla	L 550 (node 19)	Valentine 1992
	Oldest known undisputable metazoans	L 580 (node 16)	Li et al. 1998 ; Signor & Lipps 1992
	Oldest known ascomycete fungi (Pezizomycotina)	L 400 (node 13)	Taylor et al. 1999
	Tappania interpreted as a higher fungus	L 1400 (node 9)	Butterfield 2005
Plantae	Oldest known monocotyledons	L 110 (node 36)	Friis et al. 2004
	Oldest known Nymphaeales	L 115 (node 35)	Friis et al. 2001
	Oldest known coniferophytes	L 290 (node 34)	Soltis et al. 2002
	Oldest known spermatophytes	L 380 (node 33)	Soltis et al. 2002
	Oldest known tracheophytes	L 425 (node 32)	Kenrick & Crane 1997
	Oldest known spores of land plants	L 475 (node 31)	Gray 1993
	Oldest known Florideophyceae (sister-group of Bangiales)	L 550 (node 26)	Xiao et al. 1998
	Proterocladus interpreted as a cladophoracean green alga	L 750 (node 30)	Butterfield et al. 1994
	Bangiomorpha interpreted as a red alga	L 1200 (node 26)	Butterfield et al. 1990 ; Butterfield 2000
Rhizaria	Oldest known forams (sister-group of *Gromia*)	L 525 (node 63)	Culver 1991, 1994
	Vase-shaped microfossils interpreted as testate, filose amoebae	L 750 (node 67)	Porter & Knoll 2000 ; Porter et al. 2003
Stramenopiles	Appearance of the raphid, pennate diatoms	L 50 (node 82)	Strel'nikova 1990
	Range of the genus *Raphoneis*	L 70 (node 81)	Kooistra & Medlin 1996
	Appearance of the rhizosolenid diatoms	U 90 (node 79)	Sinninghe Damsté et al. 2004
	Appearance of the rhizosolenid diatoms	L 90 (node 78)	Sinninghe Damsté et al. 2004
	Absence of pennate diatoms in Mid-Cretaceous	U 110 (node 81)	Kooistra & Medlin 1996
	Presence of the two main diatom clades in Mid-Cretaceous	L 115 (node 77)	Kooistra & Medlin 1996
	Oldest known diatom	L 185 (node 76)	Rothpletz 1896 ; Barron 1993
	Palaeovaucheria interpreted as a xanthophyte alga	L 1000 (node 75)	Hermann 1981, Woods et al. 1998
	Jacutianema solubila (a xanthophyte alga?)	750	Butterfield 2004

[a] Times are given in million years and node numbers refer to Figure 1; L: lower limit; U: upper limit.

Additional references not cited in the main text :

Barron, J. A. 1993 Diatoms. In *Fossil Prokaryotes and Protists* (ed. J. H. Lipps), pp.155–167. Boston: Blackwell Scientific Publications.

Brocks, J. J., Buick, R., Summons, R. E. & Logan, G. A. 2003 A reconstruction of Archean biological diversity based on molecular fossils from the 2.78 to 2.45 billion-year-old Mount Bruce Supergroup, Hamersley Basin, Western Australia. *Geochim. Cosmochim. Acta* **67**, 4321–4335.

Culver, S. J. 1991 Early Cambrian Foraminifera from West Africa. *Science* **254**, 689–691.

Culver, S. J. 1994 Early Cambrian Foraminifera from Southwestern Taoudeni Basin, West Africa. *J. Foraminiferal Res.* **24**, 191–202.

Friis, E. M., Pedersen, K. R. & Crane, P. R. 2001 Fossil evidence of water lilies (Nymphaeales) in the Early Cretaceous. *Nature* **410**, 357–360.

Friis, E. M., Pedersen, K. R. & Crane, P. R. 2004 Araceae from the Early Cretaceous of Portugal: evidence on the emergence of monocotyledons. *Proc. Natl. Acad. Sci. USA* **101**, 16565–16570.

Gray, J. 1993 Major Paleozoic land plant evolutionary bio-events. *Palaeogeogr. Plaeoclimatol. Palaeoecol.* **104**, 153–169.

Kenrick, P. & Crane, P. R. 1997 The origin and early evolution of plants on land. *Nature* **389**, 33–39.

Li, C. W., Chen, J. Y. & Hua, T. 1998 Precambrian sponges with cellular structures. *Science* **279**, 879–882.

Rothpletz, A. 1896 Über die Flysch-Fucoiden und einige andere fossile Algen, sowie über liasische Diatomeen führende Hornschwämme. *Zeitschr. Deutsche Geol. Ges.* **48**, 858–914.

Sáez, A. G., Probert, I., Geisen, M., Quinn, P., Young, J. R. & Medlin, L. K. 2003 Pseudo-cryptic speciation in coccolithophores. *Proc. Natl. Acad. Sci. USA* **100**, 7163–7168.

Signor, P. W. & Lipps, J. H. 1992 Origin and early radiation of the Metazoa. In *Origin and early evolution of the Metazoa* (ed. J. H. Lipps & P. W. Signor), pp. 3–23. New York: Plenum Press.

Soltis, P. S., Soltis, D. E., Savolainen, V., Crane, P. R. & Barraclough, T. G. 2002 Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. *Proc. Natl. Acad. Sci. USA* **99**, 4430–4435.

Strel'nikova, N. I. 1990 Evolution of the diatoms during the Cretaceous and Paleogene periods. In *Proceedings of the tenth international diatom symposium* (ed. H. Simola), pp. 195–204. Königstein: Költz Scientific Books.

Taylor, T. N., Hass, H. & Herp, H. 1999 The oldest fossil ascomycetes. *Nature* **399**, 648.

Valentine, J. W. 1992 The macroevolution of phyla. In *Origin and early evolution of the Metazoa* (ed. J. H. Lipps & P. W. Signor), pp. 525–553. New York: Plenum Press.

Xiao, S., Zhang, Y. & Knoll, A. H. 1998 Three-dimensional preservation of algae and animal embryos in a Neoproterozoic phosphorite. *Nature* **391**, 553–558.

4) Supplementary Table 2.

Dates calculated for the radiation of extant eukaryotes and selected basal internal nodes following four different hypotheses on the position of the root of the eukaryote phylogeny.

	root between unikonts and bikonts (Figure 1) (O,(A,(F,M)),(E,(P,C)))) [a]	root between anterokonts and opisthokonts (O,(F,M),(A,(E,(P,C))))	root between Amoebozoa and all other eukaryotes (O,A,((F,M),(E,(P,C))))	root between excavates and all other eukaryotes (O,E,((P,C),(A,(F,M))))
radiation of extant eukaryotes (node 1)	1126 (948;1357)	1146 (965;1376)	1224 (1015;1495)	1184 (971;1462)
radiation of Amoebozoa (node 3)	948 (755;1179)	956 (767;1185)	1167 (955;1438)	859 (686;1064)
radiation of opisthokonts (node 8)	960 (797;1165)	999 (831;1210)	982 (820;1186)	887 (750;1066)
radiation of Fungi (node 9)	798 (634;1003)	831 (656;1043)	818 (647;1023)	736 (597;914)
radiation of Metazoa (node 16)	812 (671;985)	839 (688;1027)	828 (681;1012)	764 (643;923)
radiation of excavates (node 22)	877 (677;1093)	862 (672;1080)	863 (669;1084)	1122 (896;1402)
radiation of plants (node 24)	930 (785;1117)	914 (776;1089)	920 (778;1100)	975 (821;1170)
radiation of red algae (node 25)	741 (600;929)	728 (596;906)	733 (597;916)	775 (616;975)
radiation of green algae (node 28)	697 (565;868)	688 (561;852)	690 (565;853)	723 (586;905)
radiation of the other bikonts (node 37)	908 (772;1085)	894 (761;1062)	898 (766;1070)	950 (804;1136)

[a] A: Amoebozoa; C: chromalveolates + Rhizaria; E: excavates; F: Fungi; M: Metazoa + choanoflagellates; O: outgroup; P: plants.

5) Supplementary Table 3. Dates calculated at each node of the tree presented in Fig. 1, with their 95% confidence intervals.

Node	Date	lower bound	upper bound	Node	Date	lower bound	upper bound
1	1126.38143	948.32668	1357.30491	42	46.78271	25.36119	64.12143
2	1094.12964	918.34838	1321.84204	43	804.00099	683.03932	964.21714
3	948.09588	754.56929	1179.17707	44	613.14141	493.21292	755.29153
4	875.65748	685.73422	1097.11732	45	476.80542	340.32554	641.05132
5	761.94443	562.69140	991.28999	46	160.57500	89.57244	262.23627
6	643.63056	455.48930	878.90309	47	480.43931	389.11864	591.03552
7	573.20064	392.36369	798.42618	48	423.50128	341.64620	519.50338
8	959.99222	796.78700	1164.67992	49	391.12875	282.31793	514.35538
9	798.37707	633.73042	1003.47208	50	314.77014	259.45734	379.24090
10	651.20518	477.44551	860.00478	51	241.31725	223.08155	249.73292
11	706.87240	544.49240	910.10017	52	196.74954	155.95218	235.31419
12	596.45859	447.63369	798.50737	53	77.97619	38.05097	124.52951
13	556.17490	414.37990	752.90011	54	233.14946	214.55653	247.40936
14	354.31760	218.20429	540.17179	55	186.37088	141.33975	229.10157
15	863.07079	710.66324	1052.21391	56	222.29813	204.83492	241.20991
16	812.12422	670.84595	984.51396	57	197.08374	163.85143	227.62202
17	662.42879	493.01678	856.23650	58	211.08610	200.39807	230.90970
18	733.12353	602.75610	892.92853	59	184.87060	152.24580	215.74125
19	591.98730	551.04731	696.12684	60	168.50984	141.27642	208.16902
20	381.67689	256.81523	524.17315	61	754.09117	639.84070	903.33860
21	1019.39680	863.76914	1225.53189	62	634.72128	543.32281	784.12362
22	876.58287	676.94293	1093.22039	63	614.25645	529.56408	758.90070
23	979.13852	835.08850	1169.06423	64	457.54569	349.94654	594.91620
24	930.21174	785.35044	1116.51490	65	414.79552	309.42916	548.41486
25	740.58359	600.24951	928.86030	66	375.85668	275.38022	505.71747
26	700.23806	566.11917	882.66237	67	292.11694	195.15966	416.29193
27	606.74016	461.32783	786.65839	68	209.16853	124.25798	320.42679
28	696.68722	564.57707	868.23574	69	207.02604	123.42448	315.58252
29	616.34310	465.63702	792.04617	70	566.67774	455.61998	700.07147
30	337.37414	205.30228	512.61304	71	530.87842	421.70569	661.96073
31	604.17383	493.82354	756.35937	72	383.82182	292.65029	498.86632
32	509.72079	430.94824	644.96035	73	313.44586	238.95458	409.62593
33	442.00848	382.95650	558.65515	74	272.34996	194.74165	368.93669
34	363.52002	294.17950	480.17129	75	187.26858	119.07725	274.75412
35	196.99507	129.13868	308.73267	76	255.98700	195.58401	338.29859
36	157.90694	111.89000	251.66467	77	222.25771	171.24660	290.14998
37	908.23159	772.15879	1084.63123	78	159.94973	114.25495	218.72655
38	552.33562	398.34167	733.25533	79	80.75913	61.20377	89.70790
39	208.72319	120.57051	329.27637	80	191.54299	145.52073	249.31228
40	160.36798	89.44412	260.76268	81	98.13259	76.59413	109.55132
41	104.35605	54.14382	175.55102	82	68.67217	51.17251	94.74345